

# Serial correlation and heterogeneous volatility in financial markets: beyond the LeBaron effect

Simone Bianco

*Department of Applied Science, College of William and Mary, Williamsburg, Va 23187-8795, USA*

Fulvio Corsi and Roberto Renò

*Dipartimento di Economia Politica, Università di Siena, Piazza S. Francesco 7, 53100, Siena, Italy*

We study the relation between serial correlation of financial returns and volatility at intraday level for the S&P500 stock index. At daily and weekly level, serial correlation and volatility are known to be negatively correlated (LeBaron effect). While confirming that the LeBaron effect holds also at intraday level, we go beyond it and, complementing the efficient market hypothesis (for returns) with the heterogeneous market hypothesis (for volatility), we test the impact of unexpected volatility, defined as the part of volatility which cannot be forecasted, on the presence of serial correlations in the time series. We show that unexpected volatility is instead positively correlated with intraday serial correlation.

Serial correlation of asset prices is one of the most elusive quantity of financial economics. According to the theory of efficient markets [1, 2], it should not exist at all, and when it exists it represents an anomaly of financial markets. Many economists and physicist devoted themselves to the study of stock return predictability [3, 4]. Historical returns should prevent any forecasting technique, even if it has been shown, like in [5], that the random walk hypothesis holds only weakly.

On the other hand the variance of financial returns on a fixed time interval, which is called volatility, is a highly predictable quantity [6, 7], with its probability distribution function showing fat tails [8]. The natural association of volatility to financial risk forecast and control makes its analysis paramount in Economics. To some extent, it seems obvious therefore to link volatility to returns serial correlation. If anything else, the link between volatility and serial correlations can reveal basic properties of the price formation mechanism.

A notable stylized fact on serial correlation is the LeBaron effect [9], according to which volatility is negatively correlated to serial correlation, relation that has been empirically confirmed at daily level on several markets [10]. We exploit, as in [11, 12], high-frequency information (that is, intraday data) to test this effect more precisely than in previous studies. With respect to [11, 12], not only we extend considerably our data set, but also our work is set within the framework of the Heterogeneous Market Hypothesis [13], which implies the modification of some basic assumptions on financial market dynamics due to empirical findings. The main feature of this model is the assumption that agents must be considered heterogeneous, namely, they react differently to incoming news. This implies a completely new interpretation of the concept of time in the market, as now every agent has its own dealing time and frequency. More importantly, the transactions time, in the agent framework, has to be related to short-, medium-, and long-term decisions. Thus, volatility will be itself consistently composed by a cascade of several time components. In this work, the heterogeneous market hypothesis is the basis of our method of volatility estimation, introduced for the first time in [14]. A notable advantage in modeling volatility in the context of heterogeneous markets is that it allows to reconcile the measure with some empirical findings on volatility (e.g., long-range dependence and fat tails), a feature which simple autoregressive volatility forecasting models lacks.

The purpose of this paper is then to exploit intraday measures of volatility and serial correlation to investigate the LeBaron effect at a deeper level with a very large and liquid dataset. We find that the LeBaron effect holds. However, we also refine the finding of LeBaron by showing that volatility can be splitted into two components: a predictable one and an unpredictable one. While the first is negatively correlated with serial correlation, the latter is instead positively correlated with serial correlation.

## I. METHODS AND DATA DESCRIPTION

Let us assume to have  $r_1, \dots, r_n$  intraday logarithmic returns. To quantify volatility, we construct daily *realized volatility* measures defined as the cumulative sum of squared intraday five-minutes returns [15]:

$$RV_t = \sum_{i=1}^n r_{i,t}^2 \quad (1)$$

where  $r_{i,t}$  is the  $i$ -th return in day  $t$ . Since volatility has been proved to be log-Normal [16], we use the logarithm of  $RV_t$  to obtain Normal distributions.

Measuring daily serial correlation is more subtle, and we borrow from [12] using a modified overlapped Variance Ratio. Define:

$$\hat{\mu} \equiv \frac{1}{n} \sum_{k=1}^n r_k \quad (2)$$

$$\hat{\sigma}_a^2 \equiv \frac{1}{n-1} \sum_{k=1}^n (r_k - \hat{\mu})^2 \quad (3)$$

$$\hat{\sigma}_c^2 \equiv \frac{1}{m} \sum_{k=q}^n \left( \sum_{j=k-q+1}^k r_j - q\hat{\mu} \right)^2 \quad (4)$$

where

$$m = q(n - q + 1) \left( 1 - \frac{q}{n} \right). \quad (5)$$

We define the variance ratio as follows:

$$VR(q) = \left( \frac{\hat{\sigma}_c^2}{\hat{\sigma}_a^2} \right)^\beta. \quad (6)$$

The use of the power transformation  $f(x) = x^\beta$  makes the distribution closer to a Normal one in small samples [17]. The expression of Eq. (6) is asymptotically Normal with mean 1 and defined standard deviation.  $\beta$  is given by:

$$\beta = 1 - \frac{2}{3} \frac{\left( \sum_{j=1}^{(n-1)/2} W_k(\lambda_j) \right) \left( \sum_{j=1}^{(n-1)/2} W_k^3(\lambda_j) \right)}{\left( \sum_{j=1}^{(n-1)/2} W_k^2(\lambda_j) \right)^2}, \quad (7)$$

where  $W_k$  is the Dirichlet kernel:

$$W_k(\lambda) = \frac{1}{k} \frac{\sin^2(k\lambda/2)}{\sin^2(\lambda/2)} \quad (8)$$

and  $\lambda_j = 2\pi j/n$ .

Intuitively, the variance ratio expresses the ratio of variances computed at two different frequencies, whose ratio is given by  $q$ . If there is no serial correlation in the data,  $VR(q)$  should be close to one. In presence of positive serial correlation, the variance computed at the higher frequency is lower, and  $VR(q) < 1$ . If instead there is negative serial correlation, this argument reverts and  $VR(q) > 1$ . We use  $q = 2, 3, 4, 5, 6$ . Higher values of  $q$  cannot be used without introducing possible distortions in the statistics behavior [18]. The  $VR$  measure has been shown to be correct also for heteroskedastic data generating process [5], and it is defined with overlapping observations [18]. This measure is a reliable measure of serial correlation both at daily [5] and intraday [11, 12] level.

The data set we use is one of the most liquid financial asset in the world, that is the S&P500 stock index futures from 1993 to 2007, for a total of 4344 days. We have all high-frequency information, but to avoid for microstructure effects we use a grid of 84 daily 5-minutes logarithmic returns, interpolated according to the previous-tick scheme (the price at time  $t$  is the last observed price before  $t$ ). These choices are the standard in this kind of applications.

Relation between volatility and serial correlation can simply be tested using a linear regression

$$VR(q)_t = b + c \log RV_t + \varepsilon_t. \quad (9)$$

However, the LeBaron effect tests the relation between serial correlation and *lagged* volatility, which can be tested via the regression:

$$VR(q)_t = b + c_0 \log RV_t + c_1 \log RV_{t-1} + \varepsilon_t. \quad (10)$$

If we want to add further lagged values of realized volatility, we cannot ignore the fact that volatility is well known to display long-range dependence. One effective way to accommodate for this stylized fact without resorting to the estimation burden of a long memory model is the HAR model of [14]. Following what can be termed the "Heterogeneous Market Hypothesis" of [19, 20, 21, 22, 23] which recognizes the presence of heterogeneity in traders

horizon and the asymmetric propagation of volatility cascade from long to short time periods [24], the basic idea that emerges is that heterogeneous market structure generates an heterogeneous volatility cascade. Hence, [14] proposed a stochastic additive cascade of three different realized volatility components which explain the long memory observed in the volatility as the superimposition of few processes operating on different time scales corresponding to the three typical time horizons operating in the financial market: daily, weekly and monthly. This stochastic volatility cascade leads to a simple AR-type model in the realized volatility with the feature of considering realized volatilities defined over heterogeneous time periods (the HAR model):

$$\begin{aligned} \log RV_t = & \beta_0 + \beta_{(d)} \log RV_{t-1} \\ & + \beta_{(w)} \log RV_{t-1}^{(w)} + \beta_{(m)} \log RV_{t-1}^{(m)} + \eta_t \end{aligned} \quad (11)$$

where  $\eta_t$  is a zero-mean estimation error and:

$$\log RV_{t-1}^{(w)} = \frac{1}{5} \sum_{k=1}^5 \log RV_{t-k}, \quad \log RV_{t-1}^{(m)} = \frac{1}{22} \sum_{k=1}^{22} \log RV_{t-k}. \quad (12)$$

Although the HAR model doesn't formally belong to the class of long-memory models, it generates apparent power laws and long-memory, i.e. it is able to reproduce a memory decay which is indistinguishable from that observed in the empirical data. It is now widely used in applications in financial economics [25, 26]. We estimate the HAR model with ordinary least squares, and use the estimated coefficients  $\hat{\beta}_{0,(d),(w),(m)}$  to define the predictable volatility as:

$$\sigma_{p,t} = \hat{\beta}_0 + \hat{\beta}_{(d)} \log RV_{t-1} + \hat{\beta}_{(w)} \log RV_{t-1}^{(w)} + \hat{\beta}_{(m)} \log RV_{t-1}^{(m)} \quad (13)$$

and *unexpected volatility* as the residuals of the above regression:

$$\sigma_{u,t} = \hat{\eta}_t, \quad (14)$$

allowing us to estimate the linear model:

$$VR(q)_t = b + c_u \sigma_{p,t} + c_p \sigma_{u,t} + \varepsilon_t \quad (15)$$

which fully takes into account heterogeneity, long-range dependence and heteroskedasticity of financial market volatility.

## II. RESULTS AND DISCUSSION

We estimate the regression [9] on a rolling window five years long (Figure 2, with  $q = 2, 6$ ) as well as on the whole sample (Table 1,  $q = 2, 3, 4, 5, 6$ ). Results in the first row of Figure 2 may look disappointing, since the LeBaron effect is not displayed when looking just at the correlation between realized volatility and variance ratio. Moreover, this correlation tends to be slightly positive instead of negative, especially in the first part of the sample.

However, LeBaron evidenced the negative correlation between  $r_t^2$  and  $r_t r_{t+1}$ . This is indeed what the second row of Figure 2 shows: the coefficient of lagged volatility on variance ratio is negative and significant across all the sample.

Most interestingly, when the LeBaron effect is correctly disentangled in the data, we find that contemporaneous volatility is significantly and positively correlated with the variance ratio. Hence, estimation results for equation [10] indicate a sharp difference in the relation between serial correlation and volatility: strongly positive for contemporaneous volatility and strongly negative for lagged one. Such antithetical behavior of the relation is even more puzzling considering the well known stylized fact of volatility to be highly persistent. How could we explain this result? By our heterogeneous “rotation” of the regressors, we can rewrite Equation [10] in the form [15]. This provides the separation between predictable and unexpected volatility illustrated in Figure 1. The new specification greatly helps in shedding light on this result providing a precise economic interpretation. Hence, as [12] suggested, we can now provide an explanation in term of predictable and unexpected volatility: since volatility is known to be predictable by market participants, it has a different impact with respect to its unpredictable component.

The third row of Figure 2 shows indeed that the predictable volatility, defined by means of the HAR model, is negatively correlated with the variance ratio (more with higher  $q$ ) and that the unpredictable volatility is positively correlated with the variance ratio (more with higher  $q$ ).

The full sample estimates in Table 1 corroborate this finding. We can then rephrase the LeBaron effect as follows: serial correlation is negatively correlated with the expected volatility. Moreover, we can conclude that serial correlation

is instead positively correlated with unexpected volatility, which is a previously unrecognized stylized fact of financial returns. This stylized fact suggests that the usual explanation of the LeBaron effect in terms of feedback trading [27] is at least incomplete, advocating for a broader theory on the link between volatility and the way information is spread to heterogenous market components.

---

- [1] E. Fama (1970), Efficient capital markets: a review of theory and empirical work. *J Finance*, 25:383–417.
- [2] J. Doyne Farmer and A. W. Lo (1999), Frontiers of finance: Evolution and efficient markets. *Proc Natl Acad Sci USA* 96:9991-9992.
- [3] J. Doyne Farmer, P. Patelli and I. I. Zovko (2005), The predictive power of zero intelligence in financial markets. *Proc Natl Acad Sci USA* 102:2254-2259.
- [4] S. Pincus and R. E. Kalman (2004), Irregularity, volatility, risk, and financial market time series. *Proc Natl Acad Sci USA* 101:13709-13714.
- [5] A. W. Lo and A. C. MacKinlay (1988), Stock market prices do not follow random walks: evidence from a simple specification test. *Rev Finan Stud*, 1:41–66.
- [6] R. Engle (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987-1008
- [7] T. Bollerslev (1986), Generalized Autoregressive Conditional Heteroskedasticity. *J Econometrics*, 31:307-327.
- [8] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley (1997), Correlations in economic time series. *Physica A*, 245:437-440.
- [9] B. LeBaron (1992), Some relations between volatility and serial correlations in stock market returns. *J Bus*, 65(2):199–219.
- [10] I. A. Venetis and D. Peel (2005), Non-linearity in stock index returns: the volatility and serial correlation relationship. *Econ Modelling* 22:1-19.
- [11] S. Bianco and R. Renò (2006), Dynamics of intraday serial correlation in the Italian futures market. *J Futures Markets*, 26:61–84.
- [12] S. Bianco and R. Renò (2008), Unexpected volatility and intraday serial correlation. *Forthcoming in Quant Finance*.
- [13] U. Muller, M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker (1993), Fractals and intrinsic time - a challenge to econometricians. XXXIXth International AEA Conference on Real Time Econometrics, 14-15 Oct 1994, Luxembourg.
- [14] F. Corsi. A simple long memory model of realized volatility. PhD Dissertation, 2004.
- [15] T. Andersen, T. Bollerslev, Diebold F.X. and P. Labys (2003), Modeling and forecasting realized volatility. *Econometrica*, 71:579-625.
- [16] P. Cizeau, Y. Liu, M. Meyer, C.-K. Peng, and H. E. Stanley (1997), Volatility distribution in the S&P500 stock index. *Physica A*, 245:441-445.
- [17] W. Chen and R. S. Deo (2006), The variance ratio statistic at large horizon. *Econometric Theory*, 22:206–234.
- [18] R. S. Deo and M. Richardson, M (2003) On the asymptotic power of the variance ratio test. *Econometric Theory*, 19:231–239.
- [19] M. M. Dacorogna, R. Genay, U. A. Müller, R. B. Olsen and O. V. Pictet, *An Introduction to High-Frequency Finance*. Academic Press, 2001.
- [20] T. Lux and M. Marchesi (1999), Scaling and criticality in a stochastic multi-agent model of financial market. *Nature*, 397:498–500.
- [21] B. LeBaron (2002), Short-memory traders and their impact on group learning in financial markets. *Proc Natl Acad Sci USA* 99:7201-7206.
- [22] A. Weron and R. Weron (2000), Fractal market hypothesis and two power laws. *Chaos, Solitons and Fractals* 11:289-296.
- [23] D. G. McMillan and A. E. H. Speight (2006), Volatility dynamics and heterogenous markets. *Int J Finance Econ*, 11:115-121.
- [24] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner and Y. Dodge (1996), Turbulent cascades in foreign exchange markets. *Nature*, 381:767–770.
- [25] T. Andersen, T. Bollerslev and F. X. Diebold (2007), Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. *Rev Econ Stat*, 89:701-720.
- [26] Ghysels, E., Santa-Clara, P. and R.S. Valkanov (2006), Predicting volatility: getting the most out of return data sampled at different frequencies. *J Econometrics*, 131: 59-95.
- [27] E. Sentana, and S. Wadhwani, (1992), Feedback traders and stock return autocorrelation: evidence from a century of daily data. *Econ J*, 102: 415-425.

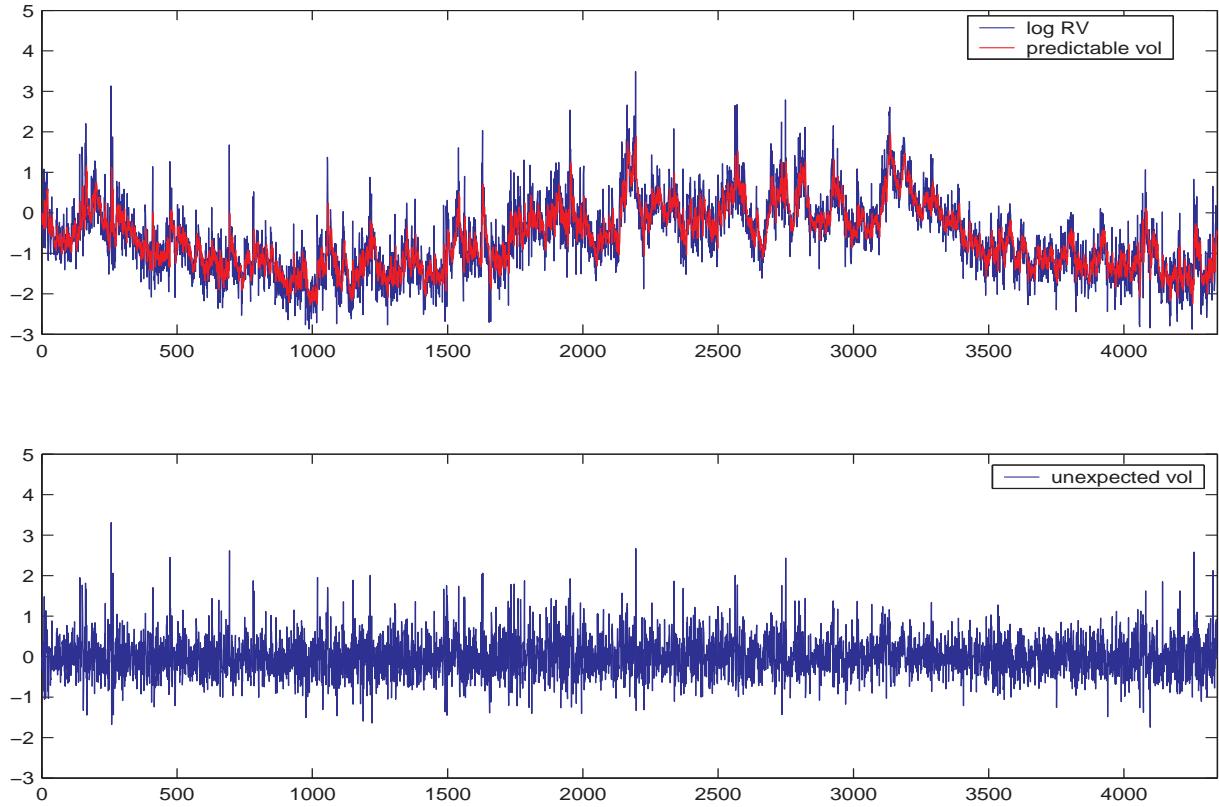


FIG. 1: Top: shows the time series estimate of  $\log RV_t$  used in this study (4344 observations), together with the predictable volatility estimated by means of the HAR model. Bottom: shows the time series of unexpected volatility estimated as the residuals of the HAR model.

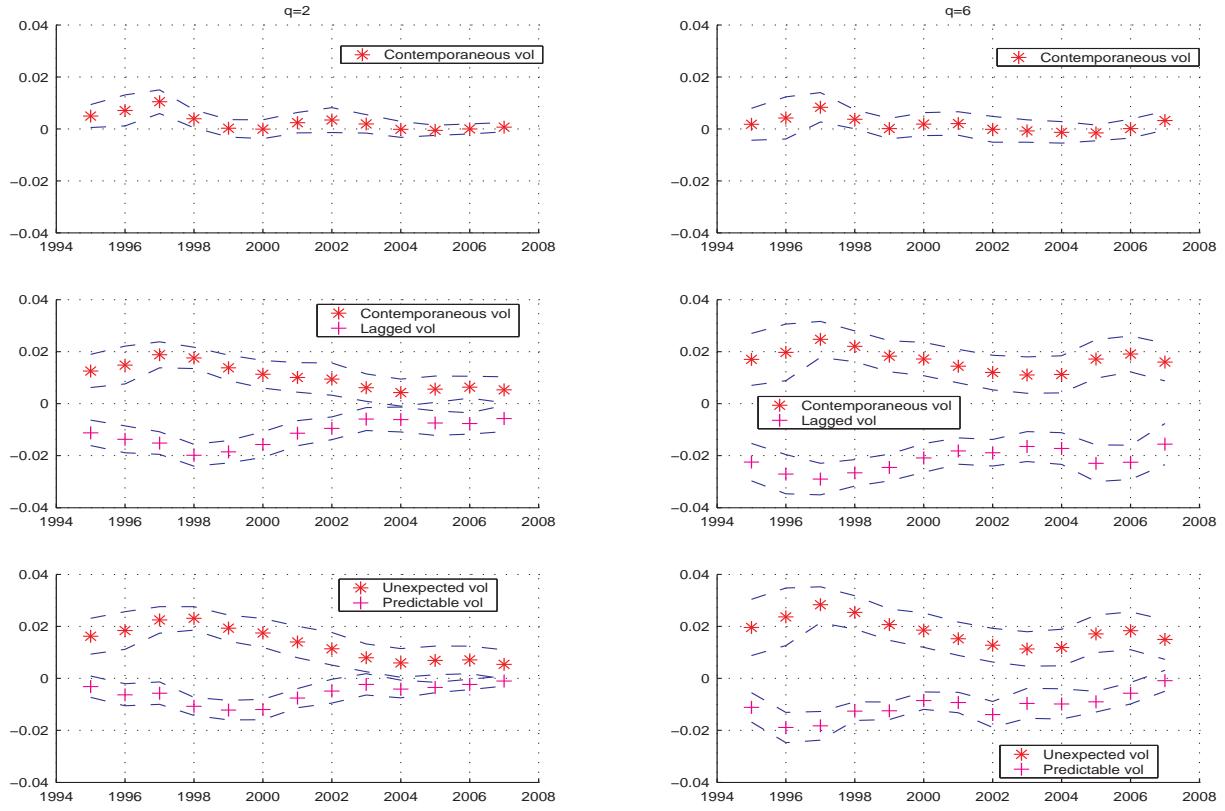


FIG. 2: On a rolling window of lenght 5 years, shows: (top) the measures of coefficient  $c_0$  of regression (9); (middle) the measures of coefficients  $c_0$  and  $c_1$  of regression (10); (bottom) the measures of coefficients  $c_0$  and  $c_1$  of regression (15). On the left column, the Variance ratio have been computed with  $q = 2$  (5 minutes); on the right column, with  $q = 6$  (25 minutes). Dashed lines represents confidence intervals at 95% confidence level. This figure shows that the correlation between serial correlation and predictable volatility is negative (LeBaron effect), while the correlation between serial correlation and unexpected volatility is positive.

TABLE I: On the full sample, estimated regression. Standard errors are in brackets. The adjusted  $R^2$  is indicated by  $\bar{R}^2$  and is expressed in percentage form.

$q$	Regression 9			Regression 10			Regression 15		
	$c$	$\bar{R}^2(\%)$	$c_0$	$c_1$	$\bar{R}^2(\%)$	$c_u$	$c_p$	$\bar{R}^2(\%)$	
2	-0.242	1.40	9.026	-11.851	3.82	12.707	-6.376	4.85	
	(0.805)		(1.506)	(1.461)		(1.702)	(0.958)		
3	-0.949	0.44	10.742	-14.950	3.59	13.478	-7.780	3.93	
	(0.933)		(1.754)	(1.656)		(1.998)	(1.156)		
4	-0.160	-0.01	12.838	-16.608	3.67	14.645	-7.080	3.45	
	(0.978)		(1.823)	(1.697)		(2.028)	(1.210)		
5	0.325	0.05	14.499	-18.108	3.95	15.860	-6.903	3.45	
	(1.051)		(1.969)	(1.777)		(2.154)	(1.205)		
6	0.810	0.23	16.477	-20.018	4.42	17.603	-6.986	3.73	
	(1.108)		(2.118)	(1.902)		(2.280)	(1.211)		